Unit 2 (Chapters 7 – 9) Review Packet

Key

Use the data below for questions 1 through 14

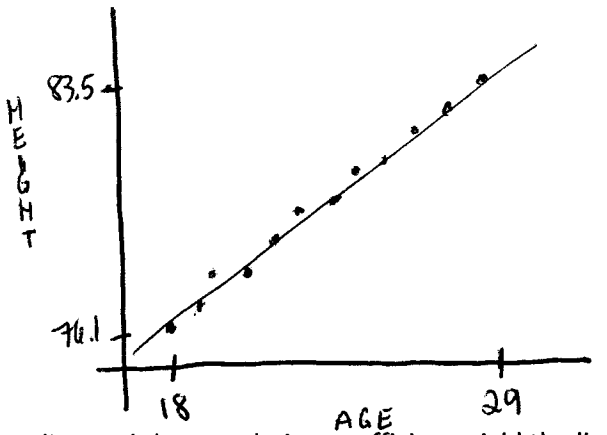
Below is data concerning the mean height of Kalama children. A scientist wanted to look at the effect that age had on the mean height of the children.

Age (months)	18	19	20	21	22	23	24	25	26	27	28	29
Height (cm)	76.1	77	78.1	78.2	78.8	79.7	79.9	81.1	81.2	81.8	82.8	83.5

1. Determine the explanatory and response variables

age height

2. Create a scatterplot of the data. Be sure to label the axes. DESCRIBE the plot.



linear positive strong

3. Find the LSR line and the correlation coefficient. Add the line to your plot.

$\hat{y} = 64.928 + 0.635x$ $r = 0.994$

4. What proportion of the variability in the height of Kalama children is explained by the variability in their age?

$r^2 = 98.876\%$

5. Interpret the slope of the LSR line in a complete sentence

For every 1 month increase in age, the height increases by 0.635 cm.

6. Predict the mean height of a child that is 42 months old (show work!). Are you confident in your prediction? Why or why not?

$\hat{y} = 64.928 + 0.635(42)$

$\hat{y} = 91.597$ cm

Not very confident – extrapolation!

7. Predict the mean height for a child who is 24 months old (show work!)

$\hat{y} = 64.928 + 0.635(24)$ $\hat{y} = 80.167$ cm

very confident in this one – it is in the x-range and there is a high r^2.

8. Find the error of your predicted value for a child who is 24 months old

$79.9 - 80.167 = -0.267$ cm

9. Was your prediction an overestimate or an underestimate?

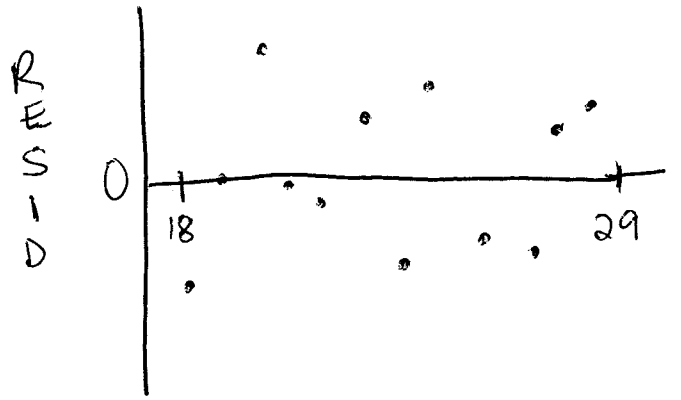
10. How old is a child expected to be if they are 100cm long? (show work!)

$100 = 64.928 + 0.635x$

$x = 55.231$ months

11. What is the sum of the residuals for this data? 0

12. Create a residual plot below. Describe the FORM of the plot.



scattered

13. What does the plot tell you about your linear model? Explain BRIEFLY.

the linear model is a good model b/c the residual plot is scattered.

14. What conclusions can be made from the previous questions? Does the age of a child CAUSE the height of the child? Why or why not?

NO! association ≠ causation. age + height are associated. our linear model does a good job of describing this relationship b/c scattered resid plot + linear scatterplot + high r.

15. Given the following data about variables x and y calculate by hand (using AP formulas) the LSR line. Show all work! Write the line in the form y = b0 + b1x .

	X	Y	
Mean	45.6	37.2	r = 0.765
St. Dev	3.2	2.1	

$b_1 = r\left(\frac{S_y}{S_x}\right) = 0.765\left(\frac{2.1}{3.2}\right) = 0.502$

$b_0 = \bar{y} - b_1\bar{x} = 37.2 - (0.502)(45.6) = 14.309$

$\hat{y} = 14.309 + 0.502x$

16. Below is a Minitab statistical analysis. The data is looking at clothes salespersons and examining the effect that the number of minutes spent with a customer has on the total dollar amount that the customer buys. In other words, if a salesperson spends more time with a customer, does the customer buy more clothing (increasing the commission of the salesperson)?

Predictor	Coeff	s.e.	T	P
Constant	-1.731	2.4065	-0.876	0.4561
Minutes	0.5679	0.00456	6.6898	1.2358

S = 1.3425 R-Sq = 0.7896 R-Sq (adj) = 0.7748

(a) What is the equation of the LSR line?

$\hat{y} = -1.731 + 0.5679x$

(b) What is the value of the correlation coefficient?

$r = \sqrt{0.7896} = 0.8886$

(c) What does the correlation tell you about the relationship of your two variables?

strong, linear, positive

(d) Interpret the slope in the context of the problem

for every 1 min spent with a customer, the total dollar amount spent goes up by $0.57.

(e) What is the coefficient of determination? Interpret this value in context of the problem.

$r^2 = 0.7896$ 78.96% of the change in $ spent is explained by the change in minutes.

(f) How much is a customer expected to buy if a salesperson spends 45 minutes with them?

$\hat{y} = -1.731 + 0.5679(45)$

$\hat{y} = \$23.82$

(g) A salesperson spent 35 minutes with a customer and the total sale was $78.50. What is the residual?

$\hat{y} = -1.731 + 0.5679(35)$

$\hat{y} = \$18.15$

resid = 78.50 − 18.15

= $60.35

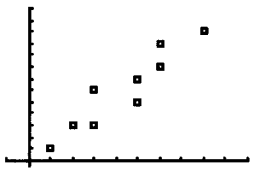
17. What does a residual plot tell us? What do we look for in a residual plot?

whether our linear model is a good model (can also show lurking variables). We are looking for a scattered residual plot to say that our linear model is a good model.

18. What type of relationship does r measure?

Linear!

19. For the graph below, what would be the closest approximation to the correlation coefficient?



(a) 0.2 (b) 0.88 (c) ~~-0.9~~ (d) ~~-0.2~~ (e) ~~0~~ (f) 0.5

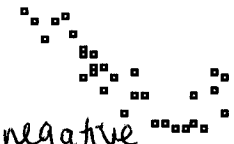
20. What is the difference between outliers, influential points, and high leverage points?

outliers - any deviation from the overall pattern of data,

Influential - strongly change slope when removed.

high leverage - outlier in the x direction

24. Describe the following plots:

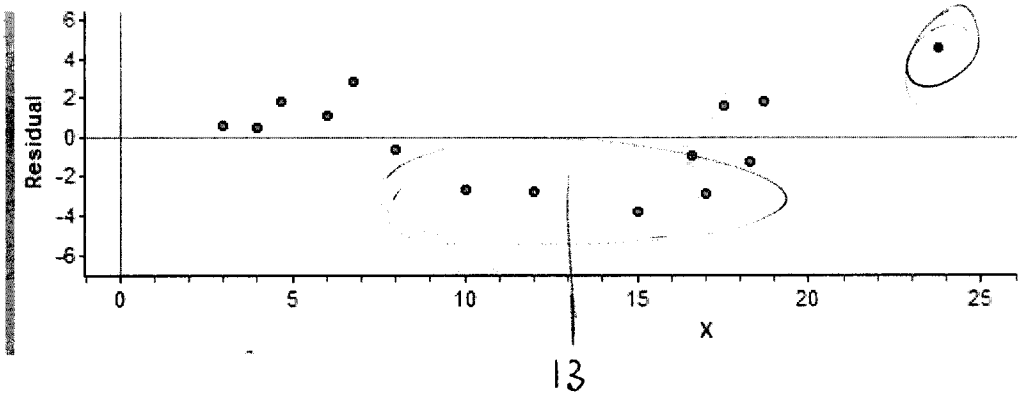


(a) negative
curved
mod. strong

(b) 2 clusters
positive

(c) curved
strong
negative, then positive

25. Look at the following residual plot.



(a) Would you expect the model to overestimate or underestimate for a prediction from an x-value of 13? Explain.

@ 13, we'd expect a negative residual, so the model overestimated

(b) Are there any outliers, high leverage points, or influential points? Identify any, and tell what type of point it is.

≈ (25, 5) - outlier, high leverage, influential